nature portfolio

Peer Review File

Mapping genetic regulation of gene expression to cellular contexts identifies long non-coding RNAs associated with brain disorders

Corresponding Author: Professor Tianyi Zhao

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscripts describes an interaction eQTL analysis using data from dorsolateral prefrontal cortex samples. While they focus on long non-coding RNAs, the conclusions also contain the results for protein-coding genes. The key point of the paper is that cell type heterogeneity of the data is resolved by deconvolution. The paper is well-written and the statistical analysis carefully executed.

1.) I was not able to access the Brain-ieQTLAtlas at the URL shown in the Discussion Because of the word wrapping in the text, I wasn't sure if the correct URL is http://brain.hitxqtl.org.cn/Brain-ieQTLAtlas/#/ or http://brain.hitxqtl.org.cn/BrainieQTLAtlas/#/ ; I tried both but neither worked. Also, you may want to use an https URL.

2.) The paper focuses on lncRNAs and mentions them in the title. However, protein-coding genes are also included in the results, and I didn't find any striking difference in the results for protein-coding genes and lncRNAs. I agree that deconvolution methods are most useful for lncRNAs because of their low expression levels, making them difficult to detect in single-cell expression data. Still, the results for protein-coding genes are also valuable, and if there are no striking differences in the results for protein-coding and non-coding genes, it may be better to put less emphasis on long non-coding RNAs. Otherwise, the reader may expect some conclusions specifically for long non-coding RNAs.

3.) The paragraph on mashr was a bit difficult to follow, and may benefit with a sentence explaining mashr and its purpose.

4.) The sentence at lines 152-155 (This pattern of cross-cell-type sharing ... correlated or anticorrelated cellular contexts) was also a bit hard to follow.

5.) Why was the analysis restricted to DLPFC? I assume the same analysis could have been applied to other brain tissues also.

6.) In line 84-85: How are "common genetic variants" defined?

7.) The deconvolution benchmarking analysis uses the correlation and the mean absolute error for its evaluation. I am happy to see the mean absolute error included in the analysis, as the correlation is highly affected by the range of expression between genes, and can give very misleading results. Still, I am wondering what the correlation and MAE values are if you compare different cell types. Do you then find a much lower correlation / much higher MAE? Maybe a heatmap of correlation values or MAE values would help, showing high correlations / low MAE values on the diagonal only.

8.) The term "genetic regulation" confused me at first, as I wasn't sure if the manuscript is about regulation by lncRNAs or regulation of lncRNAs.

9.) Please fix the author for reference 10 (by Consortium. T. Gte.).

10.) In Figure 3d, why are the columns for 5'-UTR and 3'-UTR not empty for lncRNAs?

(Remarks on code availability)
Whereas the code is provided on github, there is no explanation on how to install or run the code.

Reviewer #2

(Remarks to the Author)
This manuscript presents a large-scale interaction eQTL (ieQTL) resource for lncRNAs in the human dorsolateral prefrontal cortex (DLPFC), leveraging computational deconvolution of bulk RNA-seq data from 2,443 individuals across seven cohorts. The authors identify 3,763 ieLncRNAs with cell-type–dependent genetic regulation, integrate these with brain-related GWAS data, and demonstrate that a substantial fraction of colocalization signals are missed by bulk eQTL analyses. The work is timely, technically solid, and addresses a genuine gap in the field. I have several major and minor concerns.
1. The deconvolution evaluation was conducted using simulated or pseudobulked data. How do the methods perform on cortex data with matched bulk and single-cell data? (See: https://doi.org/10.1371/journal.pcbi.1008120)
2. For the FACS-sorted data deconvolution, does the expression of non-target cell types reflect algorithm performance or bias introduced by antibody sorting?
3. Does gene type (protein-coding gene or lncRNA) affect cell-type deconvolution?
4. (Lines 151–152) Why was LD $R^2 > 0.2$ used as the criterion for ieQTL sharing instead of π_1?
5. (Line 171) What does "well-defined" effect mean in this context? It is also unclear whether Figure 3b represents results from independent tests for each cell-type proportion or from a joint model.
6. (Line 172) The authors claim that 74% of ieLncRNAs are novel. However, the current paper provides limited independent validation of these novel transcripts.
7. (Lines 181–182) The authors claim that 85.4% of ieGenes have not been reported in previous studies and show that fewer than half of previously reported signals are replicated. Is this discrepancy fully explained by the inclusion of lncRNAs?
8. The replication of ieQTLs in single-cell data (Figure 3c), is it driven by protein-coding genes or lncRNAs? What are the replication rates for each category specifically?
9. The replication of ieQTLs in Fujita et al. and Haglund et al. shows notably different patterns, particularly for Ast. What accounts for this discrepancy? Furthermore, based on the replication results, it appears challenging to robustly define "cell-type-specific" ieQTLs. For example, for Ex and In neurons, where signals appear broadly shared across cell types. How do the authors justify the cell-type-specificity claims given these replication patterns?
10. (Lines 233–234) The authors state that "overall, ieQTLs were enriched within cell-lineage–specific open chromatin regions." However, Figure 4a shows no cell-type specificity in NeuN+ peaks, a finding further supported by the subsequent scATAC-seq data.
11. ieQTLs are shared across neuronal subtypes, yet the enrichment of ATAC-seq peaks is not subtype-specific in neuronal ieQTLs. What is the biological interpretation of this discrepancy?
12. The authors correctly identify that negative-direction ieQTLs show poor replication in matched cell types and are more susceptible to confounding from correlated cell-type proportions. However, these signals are not simply excluded and they are retained in the count of 3,763 ieLncRNAs and contribute to downstream analyses. The manuscript should more explicitly state what fraction of the final ieLncRNA catalog consists of negative-direction ieQTLs, and whether the GWAS colocalization results are driven predominantly by positive-direction signals. If negative-direction ieQTLs are unreliable, their inclusion inflates the headline numbers.
13. The colocalization analyses use FDR < 0.25 for ieQTL selection, justified by reduced statistical power in interaction mapping. While this is precedented, it substantially increases the false positive rate among the input ieQTLs. The authors should provide a sensitivity analysis showing how colocalization results change under stricter thresholds (e.g., FDR < 0.05, FDR < 0.10), and explicitly discuss what proportion of the 142 colocalizing lncRNAs might be spurious.

Minor comments
1. Line 85, full name of PCG is missing.
2. Ref 37 and ref 38 are duplicated.
3. Ref 64 is the Slingshot paper for trajectory inference. This appears to be a citation error for the qvalue package.

(Remarks on code availability)
Author well documented the code for ieQTL mapping with good annotations. But I did not see any code for lncRNA mapping.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have adequately addressed the concerns I raised in my report.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The author has addressed my questions. Following minor issues need to be corrected.
1. EPIGABA is listed as syn12034263 in Methods (line 449) but syn4588488 in Data Availability (line 630).
2. Reference 51 (Ghafouri-Fard et al.) is marked "RETRACTED" in the reference list but is still cited in the main text (line 280) to support the functional relevance of MIR124-3. This citation should be replaced with a non-retracted source.
3. Methods (line 467–468) cites "Emani et al.45" for cell-type-specific chromatin accessibility peaks, but reference 45 is Nott et al.

(Remarks on code availability)
I was able to install and run the code following the instructions provided in the README file.

Reviewer's Comments:

We sincerely thank both reviewers for their positive assessment of our study and for their constructive comments, which have helped us significantly improve the manuscript. In response to these comments, we have performed additional analyses, revised the interpretation of key results, and improved the clarity and accessibility of the associated resources and code. We have addressed all the reviewers' comments point-by-point below (our responses are in **blue**) in this document and made the corresponding changes in the manuscript files (highlighted in **red**).

Summary of major additions to the revised manuscript:

1. We strengthened deconvolution benchmarking by adding cross-cell-type evaluation, ROSMAP IHC-based validation, and gene-set sensitivity analyses.
2. We revised unclear text throughout the manuscript to better define the study scope, clarify the role of lncRNAs and PCGs, and improve the interpretation of ieQTL sharing, mashr-based analysis, and positive- versus negative-direction ieQTLs.
3. We added gene-category-stratified replication analyses and assessed GWAS colocalization robustness under more stringent ieQTL thresholds.
4. We improved resource accessibility and reproducibility by correcting the Brain-ieQTLAtlas URL, expanding GitHub documentation, and adding lncRNA mapping code.

Reviewer #1 (Remarks to the Author)

This manuscripts describes an interaction eQTL analysis using data from dorsolateral prefrontal cortex samples. While they focus on long non-coding RNAs, the conclusions also contain the results for protein-coding genes. The key point of the paper is that cell type heterogeneity of the data is resolved by deconvolution. The paper is well-written and the statistical analysis carefully executed.

Re: We thank the reviewer for the summary of our study and the positive comments. We also appreciate the reviewer's constructive suggestions, which have helped us improve the clarity and presentation of our manuscript. Guided by these comments, we have revised the manuscript and added additional analyses to better clarify the study focus, deconvolution benchmarking, interpretation of ieQTL results, and accessibility of the associated resource and code.

1.) I was not able to access the Brain-ieQTLAtlas at the URL shown in the Discussion Because of the word wrapping in the text, I wasn't sure if the correct URL is http://brain.hitxqtl.org.cn/Brain-ieQTLAtlas/#/ or http://brain.hitxqtl.org.cn/BrainieQTLAtlas/#/; I tried both but neither worked. Also, you may want to use an https URL.

Re: We thank the reviewer for pointing out the issue with accessing the Brain-ieQTLAtlas. We apologize for the confusion caused by the formatting of the URL in the Discussion section. We have now corrected the URL and ensured that it is presented in a properly formatted and clickable form using a secure https link (**Figure R1 below**). The updated link is:

https://brain.hitxqtl.org.cn/Brain-ieQTLAtlas/#/

We have also revised the manuscript to avoid line wrapping issues that may disrupt the URL and have verified that the link is fully functional (line 70 and line 351).

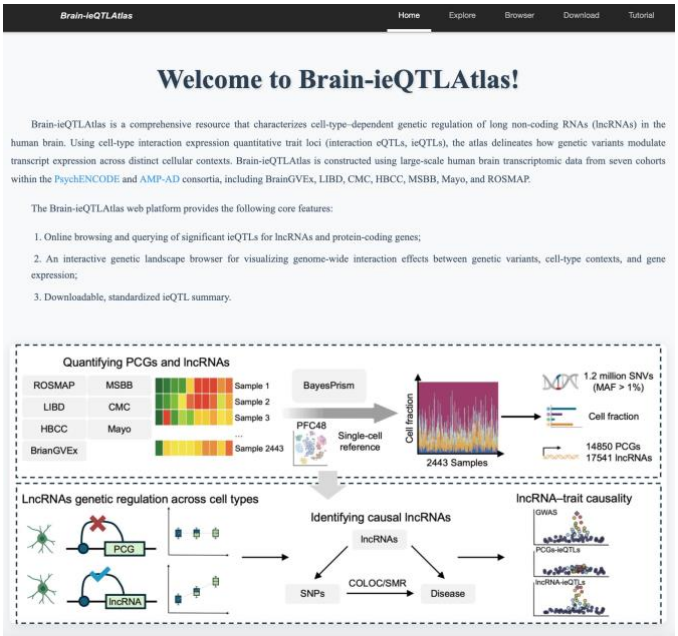

Figure R1. Overview of the Brain-ieQTLAtlas web portal interface.

2.) The paper focuses on lncRNAs and mentions them in the title. However, protein-coding genes are also included in the results, and I didn't find any striking difference in the results for protein-coding genes and lncRNAs. I agree that deconvolution methods are most useful for lncRNAs because of their low expression levels, making them difficult to detect in single-cell expression data. Still, the results for protein-coding genes are also valuable, and if there are no striking differences in the results for protein-coding and non-coding genes, it may be better to put less emphasis on long non-coding RNAs. Otherwise, the reader may expect some conclusions specifically for long non-coding RNAs.

Re: We thank the reviewer for this thoughtful suggestion. In our study, PCGs were included as an important comparative reference because extensive single-nucleus eQTL resources are available for protein-coding genes, allowing us to assess replication in matched cellular contexts and thereby evaluate the reliability of the analytical framework. PCGs also provided a reference for interpreting the relative features of lncRNA-ieQTLs. In response, we have revised the title and Abstract to better reflect the scope of the study. Specifically, the revised title now emphasizes mapping genetic regulation across cellular contexts, while positioning lncRNAs as a major biological focus and key finding rather than the sole scope of the study. In the Abstract, we now explicitly state that we performed transcriptome-wide interaction eQTL (ieQTL) mapping and then highlight the lncRNA-related discoveries as one of the principal results. We also have further clarified this role of PCGs in the revised manuscript (**lines 83–85**).

The motivation of this study stems from the systematic underrepresentation of cell-context-dependent genetic regulation of lncRNA expression in existing resources. By combining bulk RNA-seq, cell-type deconvolution, and an expanded transcriptome annotation, our study was designed to characterize this underrepresented regulatory layer while using PCG-ieQTLs as a reference within the same analytical framework.

Through this comparison, we observed that lncRNA-ieQTLs were not entirely equivalent to PCG-ieQTLs. lncRNA-ieQTLs were located closer to transcription start sites (**Figure R2 below**) and showed a more compact cis-regulatory architecture (**Figure R3 below**), suggesting relatively more localized regulatory structure. In addition, for many intronic and antisense lncRNAs, the corresponding host PCGs did not show significant ieQTL signals in the same cellular context (**Figure R4 below**), suggesting that these lncRNA regulatory signals do not always simply reflect cell-context-dependent genetic regulation of their host protein-coding genes.

GWAS integration further illustrated the added value of lncRNA-ieQTLs by enabling direct comparison between lncRNA- and PCG-ieQTL evidence at the same disease-associated loci. For example, at the schizophrenia risk locus rs12293670, previous studies suggested that this region may act through effects on the protein-coding gene NRGN. In contrast, our analysis showed that the GWAS signal colocalized with the ieQTL for the unannotated antisense lncRNA NRGN-AS1 (PP.H4 = 0.95), whereas no colocalization support was observed for the NRGN ieQTL (PP.H4 = 0.00) (**Figure R5 below**). This result refines the candidate regulatory interpretation at this risk locus from NRGN to the previously undercharacterized antisense lncRNA NRGN-AS1, illustrating how lncRNA-ieQTLs can provide more fine-grained non-coding regulatory insight into known disease-associated regions.

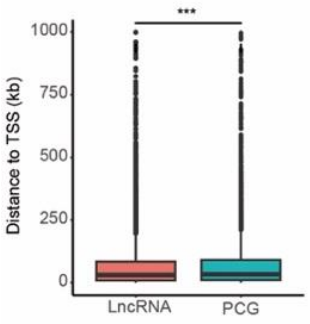

Figure R2. Distribution of genomic distances between lncRNA-ieQTLs and PCG-ieQTLs relative to the TSS. lncRNA-ieQTLs are located significantly closer to the TSS than PCG-ieQTLs. Statistical significance was assessed using a two-sided Wilcoxon rank-sum test. *p < 0.05; **p < 0.01; ***p < 0.001.

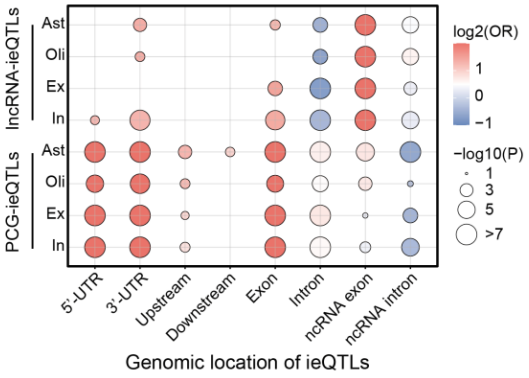

Figure R3. Enrichment analyses comparing the distribution of lncRNA-ieQTLs and PCG-ieQTLs across genomic location categories. The circle color represents log_2 odds ratio, and size indicates $-log_{10}\,P$, derived from two-sided Fisher's exact tests.

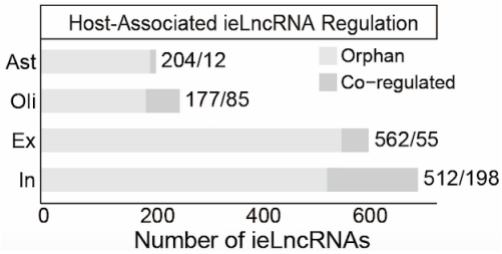

Figure R4. Among intronic and antisense lncRNAs with significant positive-direction ieQTLs, the fraction whose host protein-coding genes are also identified as ieGenes in the same cell type.

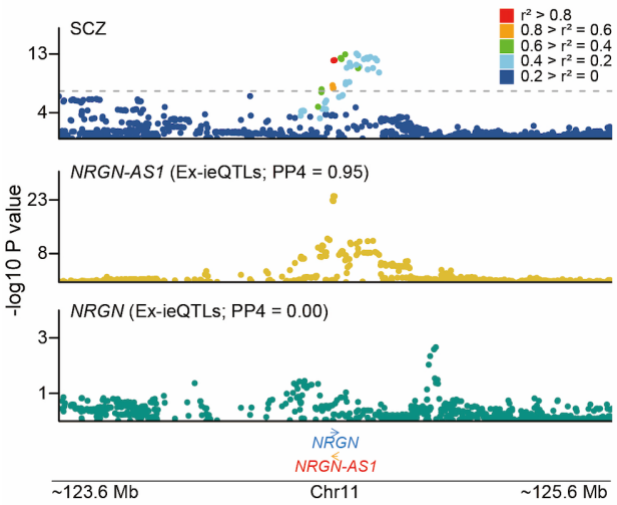

Figure R5. Locus plot showing colocalization between the SCZ GWAS locus rs12293670, NRGN, and the unannotated antisense lncRNA NRGN-AS1 in excitatory neuron ieQTLs.

Updated texts:

(lines 1-2): "Mapping genetic regulation of gene expression to cellular contexts identifies long non-coding RNAs associated with brain disorders"

(lines 20-31): "How genetic variation regulates long non-coding RNA (lncRNA) expression across brain cell types remains poorly understood. A major barrier is the resolution–power trade-off between bulk and single-nucleus expression quantitative trait locus (eQTL) studies. Here, we quantified gene expression from dorsolateral prefrontal cortex (DLPFC) RNA-seq of 2,443 individuals using an expanded transcriptome annotation and performed transcriptome-wide interaction eQTL (ieQTL) mapping with deconvolution-derived cell-type proportions. Among 17,541 analyzed lncRNAs, we identified 3,763 lncRNAs with cellular-context-dependent effects (ieLncRNAs), of which 2,783 (74%) were not annotated by GENCODE. Colocalization with genome-wide association studies (GWAS) of brain-related traits identified 118 lncRNA–trait colocalization events, approximately two-thirds of which were detectable only after modeling cellular context. Our study establishes a map of cellular-context-dependent genetic regulation in the human brain, providing a basis for nominating lncRNAs with potential roles in mediating genetic risk for brain disorders."

(lines 83-85): "We then used the expanded transcriptome model to quantify 17,541 lncRNAs, along with 14,850 protein-coding genes (PCGs) used to support cell-type deconvolution and comparative analyses."

3.) The paragraph on mashr was a bit difficult to follow, and may benefit with a sentence explaining mashr and its purpose.

Re: We thank the reviewer for this helpful suggestion. mashr[1] is an empirical Bayes framework for jointly estimating and testing effect sizes across multiple conditions. In our study, we used mashr not to identify ieQTLs de novo, but to analyze the ieQTL effect estimates across cell types and to determine whether regulatory effects were shared across multiple cellular contexts or showed evidence of cell-type specificity. Specifically, mashr learns patterns of effect-size

covariance and sharing across conditions from the data and uses these patterns to obtain adaptively shrunk posterior effect estimates. By jointly modeling ieQTL effect sizes across cell types, mashr provides posterior estimates that allow us to quantify effect sharing and cell-type specificity.
In the revised manuscript, we added a sentence explaining the purpose of mashr, clarifying that it was used to characterize the cell-type specificity and sharing patterns of ieQTLs by analyzing effect sizes across multiple conditions (**lines 230–232**). We also revised the paragraph describing the mashr-based sharing analysis to improve clarity and logical flow (**lines 232-245**) and expanded the Methods section to further describe the mashr analysis, including the input data and the construction of the strong-signal and random background sets (**lines 572–584**).

Updated texts
(lines 230-245): "To characterize cell-type specificity and sharing patterns of ieQTLs, we used multivariate adaptive shrinkage (mashr), a statistical method for analyzing effect sizes across multiple conditions[1–3] (Methods). We identified 142 lncRNA-ieQTLs and 139 PCG-ieQTLs with cell-type-specific effects across the four major brain cell types analyzed, defined by a local false sign rate[4] (LFSR) < 0.05 in only one cell type. The cell-type-specific lncRNA-ieQTLs had effect sizes comparable to those of cell-type-specific PCG-ieQTLs (**Supplementary Fig. 12c**). LncRNA-ieQTL effect sizes decreased with increasing cross-cell-type sharing, consistent with the trend observed for PCG-ieQTLs (Kruskal–Wallis test; PCG-ieQTLs, $P = 1.3 \times 10^{-10}$; lncRNA-ieQTLs, $P = 5.6 \times 10^{-15}$; **Supplementary Fig. 12d**). The inverse relationship between effect size and cross-cell-type sharing remained consistent after restricting the analysis to ieQTLs passing a more stringent significance threshold ($P_{ieQTL} < 5 \times 10^{-8}$; **Supplementary Fig. 12e**). Pairwise sharing was defined as effects significant in both cell types with consistent effect direction and mashr-estimated effect sizes within a twofold range[1]. Both lncRNA-ieQTLs and PCG-ieQTLs showed high sharing between astrocytes and oligodendrocytes, whereas sharing between excitatory and inhibitory neurons was lower for lncRNA-ieQTLs than for PCG-ieQTLs (**Fig. 3f**)."
(lines 572-584): "To estimate cell-type specificity and sharing of ieQTL effects across the four major brain cell types, we used multivariate adaptive shrinkage implemented in mashr41, a multivariate empirical Bayes method that jointly models effect estimates across conditions to obtain posterior effect-size estimates and uncertainty measures. The input to mashr consisted of genotype-by-cell-type interaction-effect estimates and their corresponding standard errors for each SNP–gene pair in each cell type. For each gene, we defined the lead variant as the tested variant with the smallest interaction P value observed across the four cell types. The resulting lead SNP–gene pairs were used as the strong-signal set for learning data-driven covariance matrices in mashr, with rows representing SNP–gene pairs and columns representing cell types. We additionally selected 250,000 tested SNP–gene pairs at random to estimate the correlation structure among effect estimates across cell types. The fitted mashr model was then applied to significant ieQTLs to obtain posterior effect-size estimates and LFSR for each SNP–gene pair in each cell type, with LFSR < 0.05 used to define significant effects."

4.) The sentence at lines 152-155 (This pattern of cross-cell-type sharing ... correlated or

anticorrelated cellular contexts) was also a bit hard to follow.

Re: We thank the reviewer for pointing out that this explanation was unclear. To improve clarity, we have revised this section to separate the observation from its interpretation (lines 163-172). We first state the observation that negative-direction ieQTLs assigned to one cell type were often in LD with positive-direction ieQTLs assigned to other cell types. We then describe the joint model that includes genotype-by-cell-type interaction terms for all four major cell types, which was used as a sensitivity analysis to evaluate whether initially significant interaction terms remained significant after modeling all cellular contexts simultaneously. The revised text now explains that positive-direction ieQTLs showed a higher retention rate in the joint model than negative-direction ieQTLs. This result suggests that negative-direction ieQTLs may have less stable cell-type attribution under the deconvolution-based interaction framework.

Updated text (lines 163-172): "We noticed that negative-direction ieQTLs in one cell type were often in linkage disequilibrium (LD) with positive-direction ieQTLs in other cell types (**Supplementary Fig. 8a–c**). Given correlations among estimated cell-type abundances, we further fitted a joint model that included interaction terms for all four cell types simultaneously and tested whether the originally significant interaction term remained significant (**Supplementary Fig. 6d**). Under this more stringent model, positive-direction ieQTLs showed higher rates of significance retention for their originally assigned interaction terms than negative-direction ieQTLs. This suggests that negative-direction ieQTLs may have less stable cell-type attribution than positive-direction ieQTLs, potentially reflecting regulatory effects from other correlated or anticorrelated cellular contexts[5,6]."

5.) Why was the analysis restricted to DLPFC? I assume the same analysis could have been applied to other brain tissues also.

Re: The analytical framework is not restricted to the DLPFC and could in principle be applied to other brain regions or tissues when sufficiently large bulk RNA-seq cohorts and appropriate single-cell or single-nucleus reference profiles are available. In the present study, we focused on the DLPFC because it represented the largest harmonized brain RNA-seq dataset available for this analysis with matched genotype data. This was important because ieQTL mapping requires substantial sample size to detect genotype-by-cellular-context interaction effects with adequate statistical power and to minimize unstable discoveries. We agree that extending this analysis to additional brain regions would provide a more comprehensive view of cellular-context-dependent genetic regulation in the human brain. We have therefore added this point to the Discussion (**lines 391–395**).

Updated text (lines 391–395): "Furthermore, our discovery dataset consisted of DLPFC samples. Although the DLPFC is highly relevant to cognitive function and neuropsychiatric disease, extending this analysis to additional brain regions will be important for obtaining a more comprehensive understanding of genetic regulation of lncRNA expression in the human brain."

6.) In line 84-85: How are "common genetic variants" defined?

Re: In this study, "common genetic variants" were defined as variants with minor allele frequency (MAF) > 0.01. We have revised the text to make this definition explicit in the study design and data overview section (**lines 81-84**) and have also clarified the corresponding genotype quality-control criteria in the Methods (**lines 520-523**).

Updated text (lines 82-83): "… and ~12 million common genetic variants with minor allele frequency (MAF) > 0.01 (Methods)."

7.) The deconvolution benchmarking analysis uses the correlation and the mean absolute error for its evaluation. I am happy to see the mean absolute error included in the analysis, as the correlation is highly affected by the range of expression between genes, and can give very misleading results. Still, I am wondering what the correlation and MAE values are if you compare different cell types. Do you then find a much lower correlation / much higher MAE? Maybe a heatmap of correlation values or MAE values would help, showing high correlations / low MAE values on the diagonal only.

Re: We added heatmaps of cross-cell-type Spearman correlation and mean absolute error (MAE) between estimated and reference cell-type proportions for each deconvolution method (**Figure R6 below**). In these heatmaps, rows represent estimated cell types and columns represent reference cell types. For BayesPrism, the correlation heatmap showed a clear diagonal pattern, with high correlations (red) for matched cell types, whereas the MAE heatmap showed low errors (blue) on the diagonal. Other methods also showed this expected trend, but the pattern was more consistent for BayesPrism. These results support accurate and stable recovery of cell-type-specific abundance variation. MAE was calculated as $MAE = n^{-1}\sum_{i=1}^{n}|\hat{p}_i - p_i|$, where $\hat{p}_i$ and p_i denote the estimated and reference proportions in sample i, respectively. For matched cell-type pairs, this quantity directly measures the absolute prediction error for the corresponding cell type. For mismatched cell-type pairs, however, MAE reflects not only potential cell-type confusion but also differences in baseline abundance and dynamic range between the two cell types being compared. For example, excitatory neurons constitute a relatively abundant cell population in cortical tissue, whereas several non-neuronal cell types have much lower proportions; consequently, comparisons between excitatory-neuron estimates and low-abundance reference cell types can naturally yield larger absolute differences. We have added these analyses to the revised Supplementary Figures 1 and updated the Results section accordingly (**lines 104-105**).

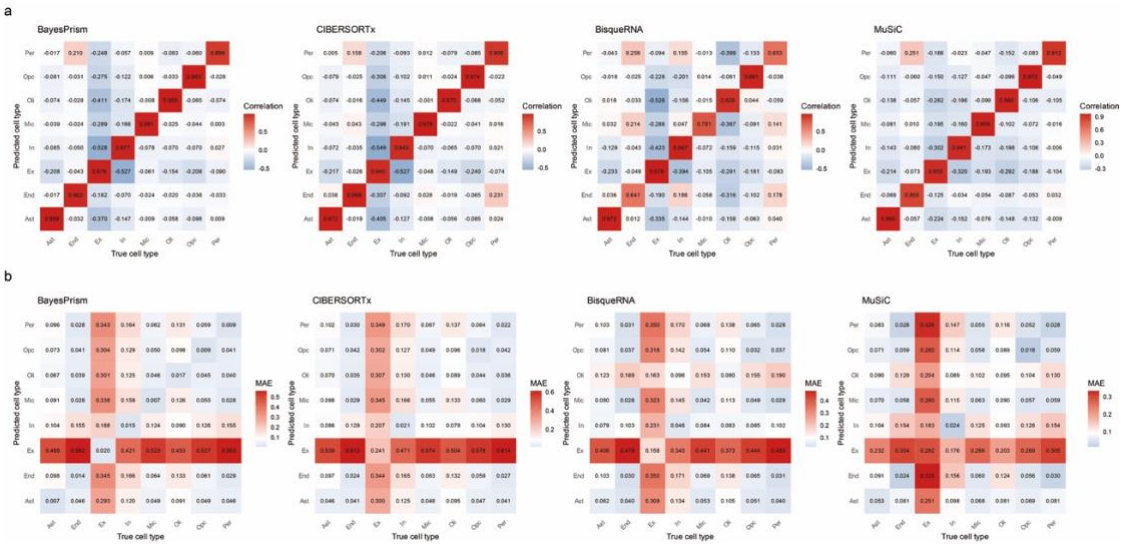

Figure R6. Cross-cell-type benchmarking of deconvolution accuracy. a. Spearman correlation heatmaps comparing estimated and reference cell-type proportions for BayesPrism, CIBERSORTx, BisqueRNA, and MuSiC. Rows represent estimated cell types and columns represent reference cell types. A clear diagonal pattern indicates that each estimated cell-type proportion most strongly tracks variation in the corresponding reference cell type. **b.** Cross-cell-type MAE heatmaps comparing estimated and reference cell-type proportions for the same methods. Abbreviations: Ast, astrocytes; End, endothelial cells; Ex, excitatory neurons; In, inhibitory neurons; Mic, microglia; Oli, oligodendrocytes; Opc, oligodendrocyte precursor cells; Per, pericytes.

Updated figure: Supplementary Figure 1
Updated text (lines 104-105): "…, with cross-cell-type benchmarking showing strongest agreement between matched estimated and reference cell types (**Supplementary Fig. 1**)."

8.) The term "genetic regulation" confused me at first, as I wasn't sure if the manuscript is about regulation by lncRNAs or regulation of lncRNAs.
Re: We thank the reviewer for pointing out this ambiguity. Our study focuses on the regulation of lncRNA expression by genetic variants, rather than regulatory effects exerted by lncRNAs. To avoid this confusion, we have revised the wording throughout the manuscript where appropriate by replacing ambiguous uses of "genetic regulation" with more specific phrases such as "genetic regulation of lncRNA expression".

9.) Please fix the author for reference 10 (by Consortium. T. Gte.).
Re: We have fixed the author information for reference 10.

10.) In Figure 3d, why are the columns for 5'-UTR and 3'-UTR not empty for lncRNAs?
Re: We apologize for the confusion. The 5'-UTR and 3'-UTR columns in Figure 3d refer to the genomic locations of ieQTLs based on ANNOVAR annotation[7], rather than to the transcript structure of the target lncRNAs themselves. Because ANNOVAR annotates variants relative to the reference gene model, an ieQTL variant associated with a lncRNA can be annotated as located in the 5'-UTR or 3'-UTR of an overlapping or nearby protein-coding

transcript. To avoid confusion, we have revised the figure legend and the corresponding text to clarify that the categories in Figure 3d represent ANNOVAR-based genomic annotations of ieQTL. The y-axis labels specify "lncRNA-ieQTLs" and "PCG-ieQTLs", and the x-axis is labeled "Genomic location of ieQTLs" (**Figure R7 below**).

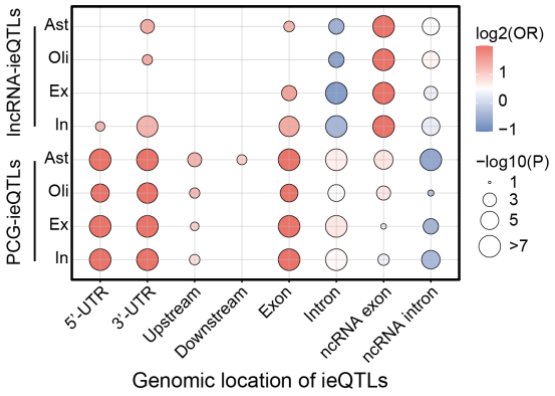

Figure R7. Enrichment analyses comparing the genomic locations of lncRNA- and PCG-ieQTLs based on ANNOVAR-defined categories[7]. Circle color represents log_2 odds ratio, and circle size indicates $-log_{10} P$, derived from two-sided Fisher's exact tests.

Updated figure: Figure 3d

Updated text (lines 862-864): "Enrichment analyses comparing the genomic locations of lncRNA- and PCG-ieQTLs based on ANNOVAR-defined categories[7]. Circle color represents log_2 odds ratio, and circle size indicates $-log_{10} P$, derived from two-sided Fisher's exact tests."

(Remarks on code availability)

Whereas the code is provided on github, there is no explanation on how to install or run the code.

Re: We have updated the GitHub repository by adding detailed instructions for code installation and execution, including the required dependencies, environment setup, and commands for running the main analysis scripts.

Reviewer #2 (Remarks to the Author)

This manuscript presents a large-scale interaction eQTL (ieQTL) resource for lncRNAs in the human dorsolateral prefrontal cortex (DLPFC), leveraging computational deconvolution of bulk RNA-seq data from 2,443 individuals across seven cohorts. The authors identify 3,763 ieLncRNAs with cell-type–dependent genetic regulation, integrate these with brain-related GWAS data, and demonstrate that a substantial fraction of colocalization signals are missed by bulk eQTL analyses. The work is timely, technically solid, and addresses a genuine gap in the field. I have several major and minor concerns.

Re: We thank the reviewer for the positive assessment of our study and for acknowledging its potential value for future studies. We are grateful for the constructive comments and suggestions, which have helped improve the clarity and rigor of the manuscript. In response, we have performed additional analyses and revised the manuscript to provide a more comprehensive assessment of data quality and methodological robustness. We address each point in detail below and hope that the revisions have adequately resolved the reviewer's concerns.

1. The deconvolution evaluation was conducted using simulated or pseudobulked data. How do the methods perform on cortex data with matched bulk and single-cell data? (See: https://doi.org/10.1371/journal.pcbi.1008120)

Re: We thank the reviewer for this helpful suggestion. We added an independent cortical benchmark using the matched ROSMAP bulk RNA-seq and immunohistochemistry (IHC) data from the study you mentioned (https://doi.org/10.1371/journal.pcbi.1008120). Using this benchmark, we found that BayesPrism achieved the highest overall concordance with IHC measurements among the evaluated methods (**Figure R8 below**). We also assessed cell-type-level correlations and error-based metrics, which further supported the robust performance of BayesPrism in ROSMAP cortical samples. These results have been added to the revised manuscript **(lines 126-132)** and are shown in Supplementary Figure 3.

We also considered the characteristics of this benchmark when interpreting the IHC comparison. As described in the original ROSMAP IHC study[8], the IHC and RNA-seq measurements were not obtained from the same tissue section, but rather from opposite hemispheres of the same donors. Therefore, comparisons between IHC-based and RNA-seq-based cell-type estimates may reflect differences in tissue location in addition to differences between experimental modalities. Moreover, IHC and bulk RNA-seq capture distinct molecular readouts, and the limited scale of the IHC experiment introduces uncertainty in the expected cell-type proportions. With these considerations in mind, the observed concordance between IHC measurements and deconvolution estimates provides additional support for the robustness of BayesPrism in real cortical bulk RNA-seq data.

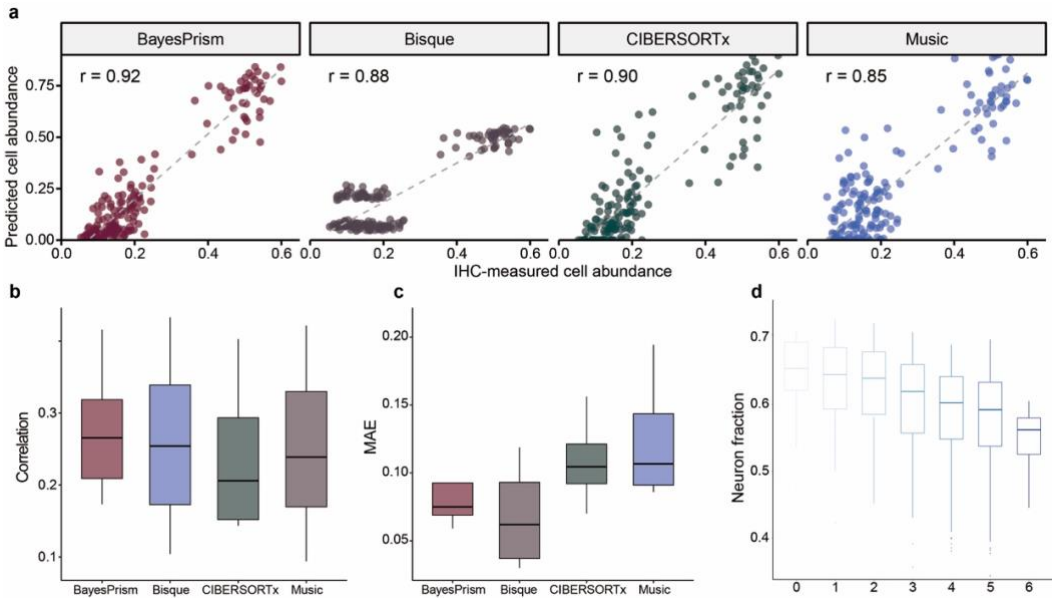

Figure R8. Benchmarking and evaluation of deconvolution estimates in ROSMAP cortical samples. a. Overall comparison of predicted and immunohistochemistry (IHC)-measured cell-type proportions in ROSMAP cortical samples (neurons, endothelial cells, microglia, oligodendrocytes and astrocytes)[8]. Each point represents one sample–cell-type pair. **b–c.** Cell-type-level benchmarking of deconvolution performance in ROSMAP cortical samples. Correlation (b) and MAE (c) were calculated separately for each cell type by comparing predicted cell-type proportions with IHC-measured cell-type proportions. IHC provides an independent reference for evaluating deconvolution estimates. However, this comparison should be interpreted with caution because the IHC data are limited in scale and because IHC and bulk RNA-seq reflect different molecular readouts[8–10]. **d.** Estimated neuronal proportions for ROSMAP cortical RNA-seq samples stratified by Braak stage. Neuronal fractions were inferred for bulk cortical RNA-seq samples from the ROSMAP cohort using BayesPrism, with the PFC48 single-nucleus RNA-seq dataset as the reference. The y-axis shows the estimated neuronal fraction inferred from bulk RNA-seq data, and the x-axis denotes Braak stages (0–6). Braak staging reflects the hierarchical progression of tau pathology in Alzheimer's disease, ranging from stage 0 (no detectable neurofibrillary tangles) to stage VI (widespread neocortical involvement)[11]. Distributions illustrate a stage-dependent decrease in inferred neuronal content with increasing Braak stage.

Updated figure: Supplementary Figure 3

Updated text (lines 126-132): "In addition, we compared deconvolution estimates with matched immunohistochemistry (IHC)-measured cell-type proportions in ROSMAP samples[8]. BayesPrism showed the highest overall concordance with IHC measurements among the evaluated methods (Spearman's rho = 0.92, compared with 0.85–0.90 for the other methods), although exact cell-type proportions remain subject to uncertainty owing to the limited scale of the IHC experiment and the distinct molecular readouts captured by IHC and RNA-seq (**Supplementary Fig. 3**)[9,10,12]."

2. For the FACS-sorted data deconvolution, does the expression of non-target cell types

reflect algorithm performance or bias introduced by antibody sorting?

Re: The FACS-sorted profiles used in our benchmark were obtained from the SuperAgerEpiMap[13] and EpiGABA[14] projects, which provide experimentally enriched human brain cell-lineage profiles with known target lineages. We agree that these samples should not be treated as fully purified ground-truth mixtures with exact cell-type proportions. We used the FACS-sorted data as a cell-type-enrichment validation setting, focusing on whether each method recovered the expected dominant lineage signal rather than interpreting residual non-target estimates as direct measures of deconvolution error. We have revised the manuscript to clarify this point (lines 118-126). In addition, the choice of BayesPrism for downstream ieQTL analyses was informed by multiple complementary benchmarks, with the FACS-enriched validation representing one component of the overall evaluation. BayesPrism showed consistently strong performance in synthetic bulk mixtures with known cell-type composition (**Figure R9 below**), cross-dataset pseudobulk profiles designed to capture inter-individual and dataset-specific heterogeneity (**Figure R10 below**), and ROSMAP cortical samples with matched IHC-derived cell-type measurements (**Figure R8 above**). These complementary results provided the basis for using BayesPrism for cell-type abundance estimation in the downstream ieQTL analyses (**lines 132-134**).

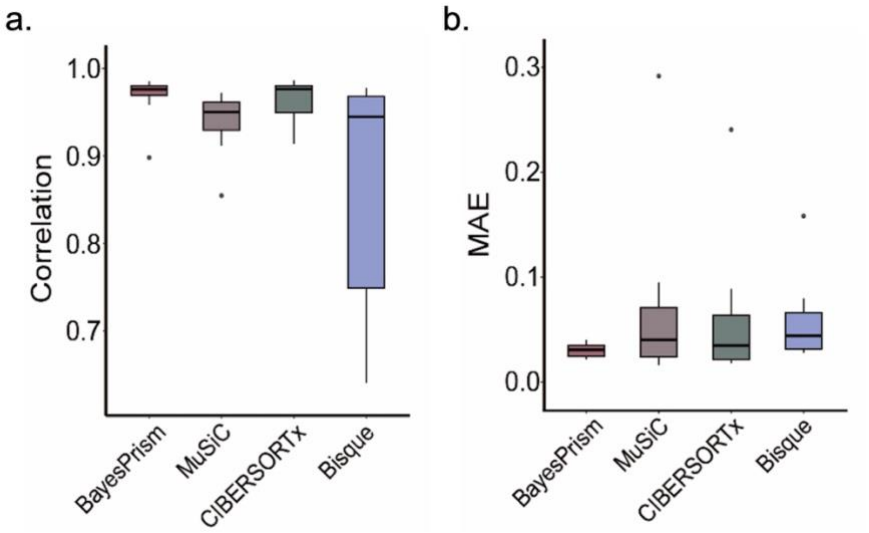

Figure R9. Evaluation of deconvolution performance. Boxplots showing the Spearman's rho (a) and mean absolute error (MAE; b) between estimated and true cell type proportions across 5,000 artificial mixtures, comparing four deconvolution methods: BayesPrism, MuSiC, CIBERSORTx, and Bisque.

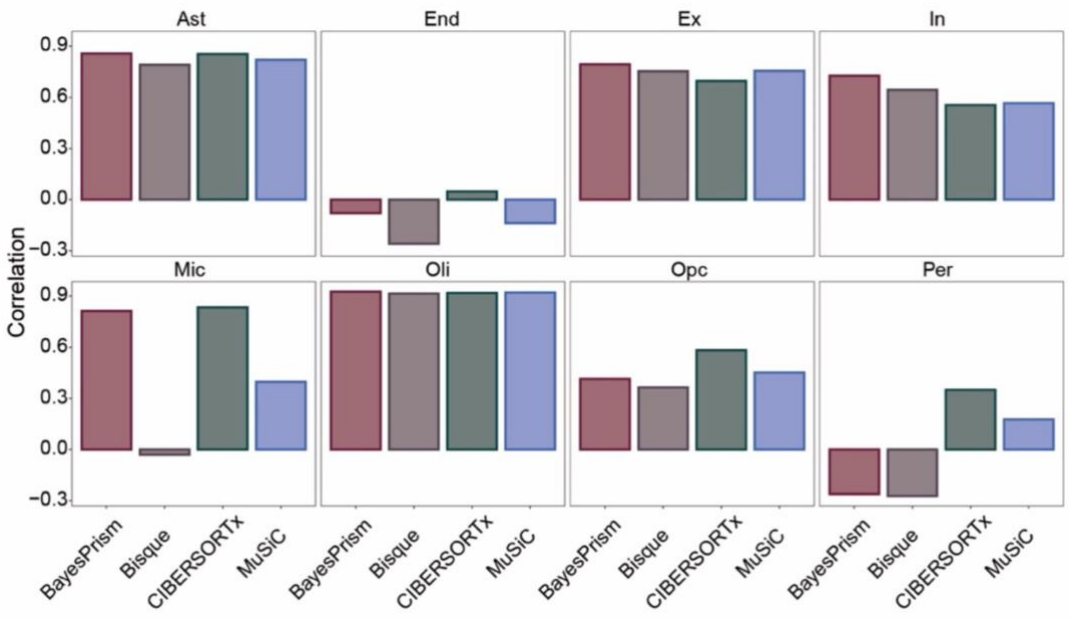

Figure R10. Benchmarking cell-type deconvolution accuracy using pseudo-bulk data. Deconvolution accuracy for eight major brain cell types evaluated on pseudo-bulk samples generated from single-nucleus RNA-seq profiles of prefrontal cortex tissue (PFC427[15], n = 427). Bars show correlations between estimated and true cell-type proportions across four methods.

Updated texts:

(lines 118-126): "This validation setting leveraged independent brain cell-type-enriched profiles with known target lineages to evaluate whether deconvolution methods recovered the expected dominant cell-type compartments. In the $NeuN^+$ and $NeuN^-$ samples (**Supplementary Fig. 2b**), BayesPrism, MuSiC, and CIBERSORTx recovered the expected dominant neuronal and non-neuronal compartments, whereas Bisque showed weaker concordance with the expected compartment structure. Similar patterns were observed in additional lineage-specific samples (**Fig. 2c**), where BayesPrism and CIBERSORTx generally aligned with the known target lineages, whereas MuSiC and Bisque showed less consistent recovery of the expected dominant cell type."

(lines 132-134): "Based on the complementary evaluations described above, we selected BayesPrism for cell-type abundance estimation in ieQTL analyses."

3. Does gene type (protein-coding gene or lncRNA) affect cell-type deconvolution?

Re: We assessed the effect of gene type on BayesPrism-based deconvolution by calculating the correlations between cell-type abundance estimates obtained using protein-coding genes only and those obtained using both protein-coding genes and lncRNAs. The estimates were nearly identical across all evaluated cell types, with correlation coefficients close to 1 (**Figure R11 below**), indicating that including lncRNAs in the input gene set had negligible impact on BayesPrism-based cell-type abundance estimation.

This observation is consistent with previous deconvolution benchmarks and method evaluations, which showed that genes with stable, detectable, and cell-type-informative expression are most useful for cell-fraction estimation, whereas lowly expressed or low-

information genes contribute little or may be filtered during preprocessing[16–19]. We have added this sensitivity analysis to the revised manuscript as Supplementary Figure 4.

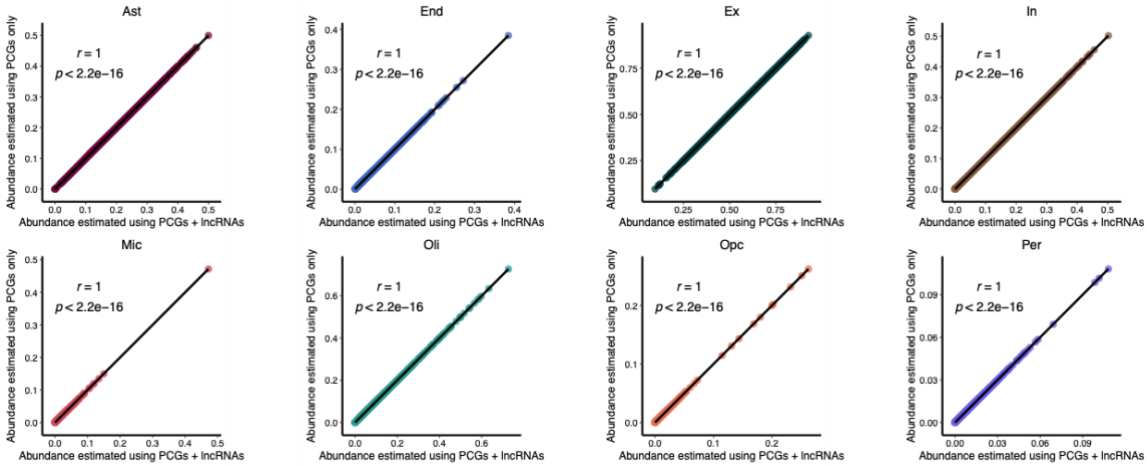

Figure R11. Comparison of BayesPrism-based cell-type abundance estimates using protein-coding genes only and using both protein-coding genes and lncRNAs. Each point represents one cortical sample. Correlation coefficients and corresponding P values are shown for each cell type. The near-identical estimates indicate that inclusion of lncRNAs had negligible impact on deconvolution results.

Updated figure: Supplementary Figure 4

Updated text (lines 140-141): "Sensitivity analysis showed that BayesPrism-based deconvolution was robust to the gene set used for abundance estimation (**Supplementary Fig. 4**)."

4. (Lines 151–152) Why was LD $R^2 > 0.2$ used as the criterion for ieQTL sharing instead of π_1?

Re: We thank the reviewer for this helpful comment. We used an LD-based criterion because ieQTLs were identified through marginal tests of SNP-by-cell-type interaction effects, and the lead variants detected across cell-type interaction models may differ while still being LD-correlated. Because interaction models yield different lead SNPs due to marginal testing noise and LD structure, direct SNP-level comparison using π_1 is not appropriate. In this setting, an LD-based criterion was better aligned with the purpose of this analysis than π_1. By contrast, π_1 is more suitable for assessing SNP-level replication of a given set of tested SNP–gene pairs across comparable analyses.

In the revised manuscript, we have expanded the legend of Supplementary Fig. 8 and the corresponding text to explicitly describe that negative-direction ieQTLs in one cell type are often in linkage disequilibrium (LD) with positive-direction ieQTLs in other cell types, and that this pattern, together with the correlation structure among estimated cell-type proportions, motivated our use of an LD-based sharing criterion and further joint modeling to assess stability of cell-type attribution (**lines 163–172**).

Updated texts

(lines 163-172): "We noticed that negative-direction ieQTLs in one cell type were often in

linkage disequilibrium (LD) with positive-direction ieQTLs in other cell types (**Supplementary Fig. 8a–c**). Together with the correlation structure among estimated cell-type abundances, this observation prompted us to further assess whether the originally assigned cell-type interaction effects remained stable in a joint model including interaction terms for all four cell types simultaneously **(Supplementary Fig. 8d**). Under this more stringent model, positive-direction ieQTLs showed higher rates of significance retention for their originally assigned interaction terms than negative-direction ieQTLs. This result suggests that negative-direction ieQTLs may have less stable cell-type attribution than positive-direction ieQTLs."

(The legend of Supplementary Fig. 8): "a–b. Pairwise LD-linked sharing of positive-direction (a) and negative-direction (b) ieQTLs across the four analyzed cell types. Sharing was quantified as the proportion of ieQTLs in a query cell type that were LD-linked to ieQTLs detected in a validation cell type, based on LD between their corresponding lead variants ($R^2 > 0.2$). Because marginal ieQTL mapping may identify different but correlated lead variants across cell-type interaction models, an LD-based criterion was used to define sharing. For each pairwise comparison, the sharing proportion was calculated as the number of LD-overlapping ieQTLs divided by the smaller number of ieQTLs detected in the two cell types. The x-axis denotes the query cell type, and the y-axis denotes the validation cell type."

5. (Line 171) What does "well-defined" effect mean in this context? It is also unclear whether Figure 3b represents results from independent tests for each cell-type proportion or from a joint model.

Re: We apologize for the unclear wording. By "well-defined" effects, we meant ieGenes with at least one positive- or negative-direction ieQTL. We have revised this wording to avoid ambiguity (**lines 180-181**). Figure 3b represents results from tests performed separately for each cell-type proportion, rather than from the joint model. We used separate tests to increase power for detecting cellular-context-dependent genetic effects. The joint model including interaction terms for all four cell types was used subsequently as a sensitivity analysis to evaluate whether the originally assigned interaction term remained significant. We have revised the Figure 3 legend to clarify this point (**lines 856-860**).

Updated texts

(lines 180-181): "We identified 3,763 interaction-effect lncRNAs (ieLncRNAs) and 3,445 PCGs (iePCGs) with at least one positive- or negative-direction ieQTL (**Supplementary Fig. 10**)."

(lines 856-860): "**b.** Number of significant ieQTLs (FDR < 0.05) identified from the interaction linear model fitted with tensorQTL[20], with tests performed separately for each of the four cell-type proportions. Results are stratified by gene class and ieQTL directionality. Bar labels indicate the total number of ieQTLs in each category."

6. (Line 172) The authors claim that 74% of ieLncRNAs are novel. However, the current paper provides limited independent validation of these novel transcripts.

Re: We thank the reviewer for raising this important point. We agree that the term "novel" could cause ambiguity, and we have clarified that the 74% value refers to ieLncRNAs that are

not annotated in current GENCODE annotations, rather than transcripts newly discovered and independently validated in the present study. These unannotated lncRNA models were derived from transcript models generated in our previous work[21]. We have revised the manuscript to replace potentially ambiguous wording such as "newly assembled lncRNAs" with "unannotated lncRNAs" or "lncRNA transcripts absent from current GENCODE annotations" where appropriate, to avoid confusion.

Importantly, the inclusion of unannotated lncRNAs allowed us to nominate additional candidate regulatory mechanisms for GWAS loci that would not be captured by analyses restricted to current GENCODE annotations. For selected unannotated lncRNAs with disease colocalization signals, we examined their cell-type expression patterns using external data. For example, Supplementary Fig. 15 shows colocalization between oligodendrocyte-related ieQTLs for the unannotated antisense lncRNA MEGF9-AS1 and educational attainment. Using the external EPIGABA dataset, we further observed oligodendrocyte-enriched expression of MEGF9-AS1 (**Figure R12 below**). In another example, Fig. 5e and Supplementary Fig. 17 show colocalization between excitatory neuron-related ieQTLs for the unannotated antisense lncRNA NRGN-AS1 and schizophrenia risk (**Figure R13 and Figure R14 below**), and external expression analysis showed enriched expression of NRGN-AS1 in excitatory neurons.

We acknowledge that these analyses do not provide complete experimental validation of all unannotated lncRNA transcript structures. They support the expression and potential disease relevance of selected unannotated lncRNAs, but direct validation of transcript structures remains beyond the scope of the present study. We have discussed this issue in the revised Discussion and clarified that unannotated lncRNAs still require continued experimental validation and careful curation (lines 389–391).

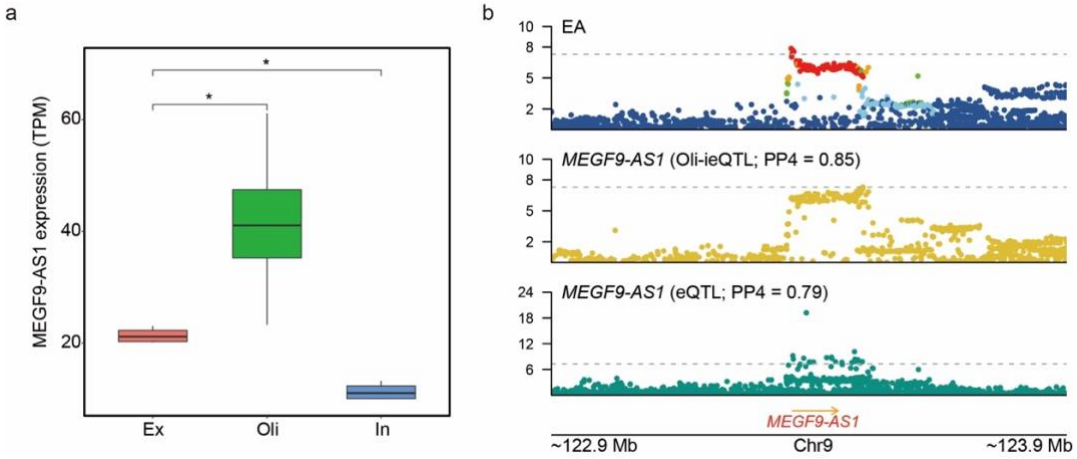

Figure R12. Colocalization of *MEGF9-AS1* with EA. a. Expression levels of the unannotated antisense lncRNA *MEGF9-AS1* (chr9: 123,348,745–123,437,964) in specific cell types: excitatory neurons (Ex), oligodendrocytes (Oli), and inhibitory neurons (In) from EPIGABA samples, highlighting its cell-type-specific expression. **b.** Locus plot showing colocalization of *MEGF9-AS1* in Oli and at the bulk-tissue level with an educational attainment (EA) GWAS signal. Association p-values ($-log_{10}$ scale) are shown for Oli-ieQTLs (middle) and bulk eQTLs (bottom). Statistical significance was assessed using a two-sided Wilcoxon rank-sum test. *p < 0.05

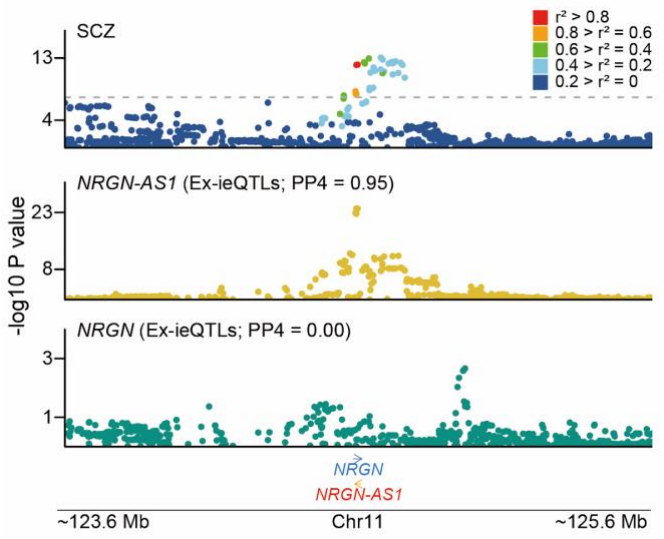

Figure R13. Colocalization of *NRGN-AS1* with SCZ. Locus plot showing colocalization between the SCZ GWAS locus rs12293670, *NRGN*, and the unannotated antisense lncRNA *NRGN-AS1* in excitatory neuron ieQTLs.

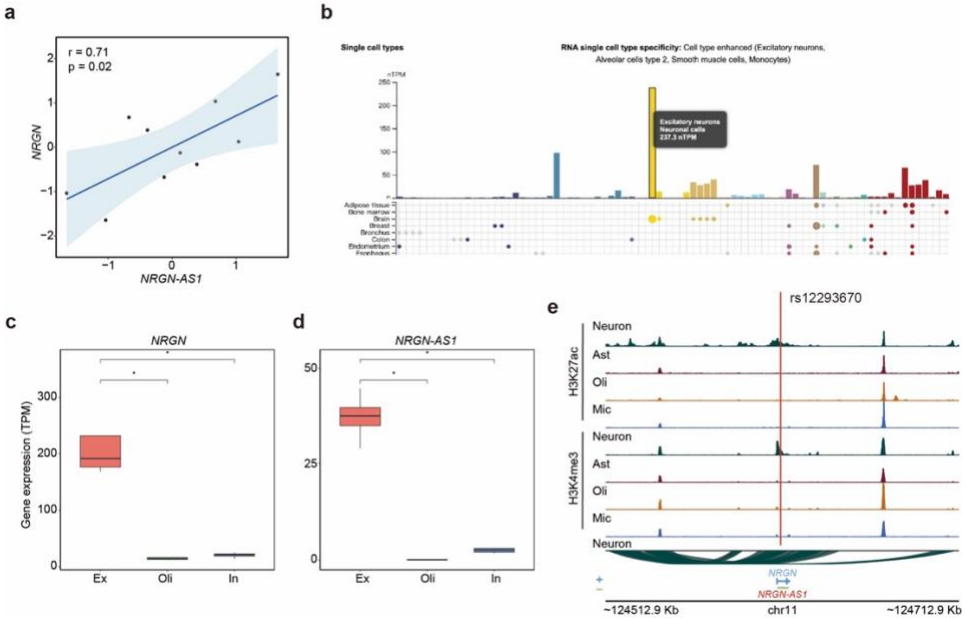

Figure R14. Molecular and epigenomic characterization of *NRGN-AS1* and *NRGN*. a. Correlation between the expression of the unannotated antisense lncRNA *NRGN-AS1* and *NRGN* in neuronal samples ($NeuN^+$). **b.** Excitatory neuron–specific expression of *NRGN* in the human brain, as reported by the Human Protein Atlas. (https://www.proteinatlas.org). **c–d.** Cell-type–specific expression of *NRGN* (c) and *NRGN-AS1* (d) across excitatory neurons (Ex), inhibitory neurons (In), and oligodendrocytes (Oli). Statistical significance was assessed using a two-sided Wilcoxon rank-sum test. *p < 0.05. **e.** Epigenomic landscape spanning the *NRGN-AS1–NRGN* locus. Tracks show cell-type–specific H3K27ac (top four tracks) and H3K4me3 (next four tracks) ChIP–seq signals in neurons, astrocytes, oligodendrocytes, and microglia, followed by neuronal PLAC-seq interaction profiles and gene annotations.

Updated text (lines 389-391): "In addition, leveraging an expanded lncRNA transcript set broadened the scope of mapping genetic regulation of lncRNA expression, but unannotated lncRNAs still require continued experimental validation and careful curation."

7. (Lines 181–182) The authors claim that 85.4% of ieGenes have not been reported in previous studies and show that fewer than half of previously reported signals are replicated. Is this discrepancy fully explained by the inclusion of lncRNAs?
Re: The inclusion of lncRNAs is an important contributor to the larger ieGene set in our study, but it is not the only factor explaining the incomplete overlap with previous studies. Several differences in study design should be considered. First, previous ieQTL studies explicitly restricted their analyses to protein-coding genes[12], whereas our analysis tested both protein-coding genes and lncRNAs, including many lncRNAs absent from current GENCODE annotations. This broader gene set increased the discovery space for identifying ieGenes. Second, the mapping strategy differed between studies. Previous studies did not perform genome-wide ieQTL mapping, which may have limited their ability to detect additional interaction signals. Third, the cellular resolution differed. Previous studies tested a larger number of cell-type contexts, whereas our ieQTL mapping focused on four major cellular contexts. This difference can affect both the discovery of new ieQTLs and the replication of previously reported signals, because some previously reported effects may be assigned to finer cell types that were not directly tested in our mapping framework.

8. The replication of ieQTLs in single-cell data (Figure 3c), is it driven by protein-coding genes or lncRNAs? What are the replication rates for each category specifically?
Re: Figure 3c reports the overall cross-cell-type replication of ieQTLs in single-cell eQTL data and was not stratified by gene type. We have now added a gene-type-stratified replication analysis for PCGs and lncRNAs (**Figure R15 below**). This added analysis was restricted to matched cellular contexts, since further stratifying the full cross-cell-type replication matrix by gene category resulted in too few tests for many cross-cell-type comparisons involving lncRNA-ieQTLs, which would lead to unstable π_1 estimates[6,22,23]. The added stratified analysis showed that the single-cell replication pattern was not driven solely by PCGs. Although PCGs generally showed more stable replication, lncRNAs also showed replication evidence in matched cellular contexts across both single-nucleus eQTL datasets, with π_1 estimates ranging from 0.61 to 0.84 in Fujita et al. and from 0.49 to 0.83 in Haglund et al. The greater variability of lncRNA replication is expected, given the lower expression levels, sparser detection, and more limited representation of lncRNAs in current resources, which limit comparable lncRNA eQTL tests and reduce the stability of π_1 estimation.

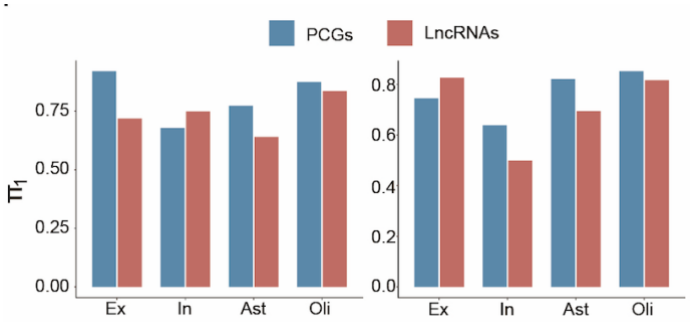

Figure R15. Gene-category-stratified replication of positive-direction ieQTLs. Bar plots show π_1 estimates for PCG-ieQTLs and lncRNA-ieQTLs in matched cellular contexts in the Fujita et al. dataset (left) and the Haglund et al. dataset (right). The greater variability of lncRNA-ieQTL replication may reflect the lower detectability and more limited representation of lncRNAs in current single-cell eQTL datasets.

Updated figure: Supplementary Figure 11
Updated text (lines 199-202): "When evaluated separately by gene category, lncRNA-ieQTLs also replicated in matched cellular contexts, although their replication estimates were slightly lower than those of PCG-ieQTLs (**Supplementary Fig. 11f**)."

9. The replication of ieQTLs in Fujita et al. and Haglund et al. shows notably different patterns, particularly for Ast. What accounts for this discrepancy? Furthermore, based on the replication results, it appears challenging to robustly define "cell-type-specific" ieQTLs. For example, for Ex and In neurons, where signals appear broadly shared across cell types. How do the authors justify the cell-type-specificity claims given these replication patterns?
Re: We thank the reviewer for raising this important point. The different replication patterns reflect differences in the degree of similarity between the two reference datasets and our discovery setting. The Fujita et al. dataset is more closely related to our discovery dataset because it was derived from ROSMAP, whereas Haglund et al. dataset represents an independent reference with differences in cohort composition, sample source, sequencing protocol, cell-type annotation, cell-state composition, and statistical power. In addition, cell-type relationships are not identical across single-cell eQTL datasets; for example, the degree of similarity or correlation between neuronal and glial subtypes can differ depending on tissue source, cell isolation, annotation strategy, and sample size. Therefore, differences in replication patterns across the two references, including for astrocyte-associated signals, are expected. Despite these differences, both reference datasets showed that ieQTLs tended to have stronger replication in the matched cellular context than in unrelated cellular contexts. This pattern supports the interpretation that the interaction effects identified in our discovery analysis capture biologically meaningful cellular-context-dependent regulatory effects.
We would like to clarify that ieQTLs are not, by definition, equivalent to strictly cell-type-specific eQTLs. Our analysis was designed to identify cell-type-dependent genetic effects in a deconvolution-based framework, where an ieQTL indicates that the genetic effect on expression is modulated by cellular context, rather than demonstrating that the effect exists exclusively in one cell type. Therefore, the sharing observed between related cell types, such as Ex and In, does not contradict the presence of cellular-context-dependent regulation. We

have revised the manuscript (**lines 344-348**) to make this distinction clearer and now consistently describe the main catalog as cellular-context-dependent ieQTLs.

Updated text (lines 361-366): "Our analysis detected many lncRNA–cell-type pairs that were not captured by single-nucleus eQTL studies, underscoring that current resources still miss a substantial fraction of cellular-context-dependent regulation of lncRNA expression. Nevertheless, deconvolution-based ieQTLs require cautious interpretation and should not be directly equated with cell-type-specific eQTLs."

10. (Lines 233–234) The authors state that "overall, ieQTLs were enriched within cell-lineage–specific open chromatin regions." However, Figure 4a shows no cell-type specificity in $NeuN^+$ peaks, a finding further supported by the subsequent scATAC-seq data.
Re: We thank the reviewer for this helpful comment. We agree that the original wording could be misleading. The purpose of Fig. 4a was to evaluate the enrichment of ieQTLs in chromatin peaks derived from $NeuN^+$ and $NeuN^-$ nuclei, with a particular focus on the difference between positive- and negative-direction ieQTLs. In the revised manuscript, we have clarified this point by emphasizing that positive-direction ieQTLs show preferential enrichment in lineage-matched peaks compared with negative-direction ieQTLs, thereby avoiding misleading interpretation of Fig. 4a **(lines 248-252).**

Updated text (lines 248-252): "To gain insight into the functional properties of ieQTLs, we annotated ieQTLs with $NeuN^+$ and $NeuN^-$ chromatin peaks identified by assay for transposase-accessible chromatin using sequencing (ATAC-seq) of FACS-sorted nuclei[24]. ieQTLs were enriched in these regulatory annotations, with positive-direction ieQTLs showing preferential enrichment in lineage-matched peaks compared with negative-direction ieQTLs (**Fig. 4a**)."

11. ieQTLs are shared across neuronal subtypes, yet the enrichment of ATAC-seq peaks is not subtype-specific in neuronal ieQTLs. What is the biological interpretation of this discrepancy?
Re: We thank the reviewer for this insightful comment. We have revised the manuscript to describe this observation more cautiously (**lines 269-270**). Our results do not support strong neuronal subtype-specific regulatory enrichment for neuronal ieQTLs. The limited enrichment in neuronal subtype-specific ATAC-seq peaks suggests that neuronal ieQTLs in our analysis may have weaker subtype specificity than glial ieQTLs. This interpretation is consistent with previous brain eQTL studies[25,26], in which neuron-associated eQTLs also showed limited enrichment in neuronal chromatin accessibility peaks.

Updated text (lines 269-270): "…, suggesting that neuronal ieQTLs may exhibit weaker cell-type specificity than glial ieQTLs, consistent with previous studies[25,26]."

12. The authors correctly identify that negative-direction ieQTLs show poor replication in matched cell types and are more susceptible to confounding from correlated cell-type proportions. However, these signals are not simply excluded and they are retained in the

count of 3,763 ieLncRNAs and contribute to downstream analyses. The manuscript should more explicitly state what fraction of the final ieLncRNA catalog consists of negative-direction ieQTLs, and whether the GWAS colocalization results are driven predominantly by positive-direction signals. If negative-direction ieQTLs are unreliable, their inclusion inflates the headline numbers.

Re: We acknowledge that negative-direction ieQTLs have less direct cell-type attribution than positive-direction ieQTLs and therefore require more cautious interpretation. However, we retained negative-direction ieQTLs because they were not devoid of supporting evidence. In our analyses, negative-direction ieQTLs also showed enrichment in functional regulatory annotations and evidence of replication in external single-nucleus eQTL resources, although their replication was less concentrated in the matched cell type. These results suggest that the main uncertainty for negative-direction ieQTLs lies in assigning them to a specific cellular context, rather than in whether they represent regulatory signals.

We have quantified the contribution of positive- and negative-direction ieQTLs to the final ieLncRNA catalog (**Figure R16 below**). This analysis showed that most ieLncRNAs had positive-direction ieQTL support: 69.6% of ieLncRNAs had at least one positive-direction ieQTL, whereas 30.4% were supported only by negative-direction ieQTLs.

We further examined whether the GWAS colocalization results were driven predominantly by negative-direction ieQTLs. In the revised colocalization result tables (Supplementary Tables 4 and 5), we added columns indicating whether each colocalized gene had positive-direction ieQTL support and the corresponding ieQTL FDR. Among the GWAS colocalization events, 64.4% had positive-direction ieQTL support, indicating that the colocalization findings were primarily supported by positive-direction signals. We have revised the manuscript to report these direction-stratified results (**lines 180-181 and lines 304-306**) and to elaborate on the interpretation of negative-direction ieQTLs in the Discussion (**lines 369-371**).

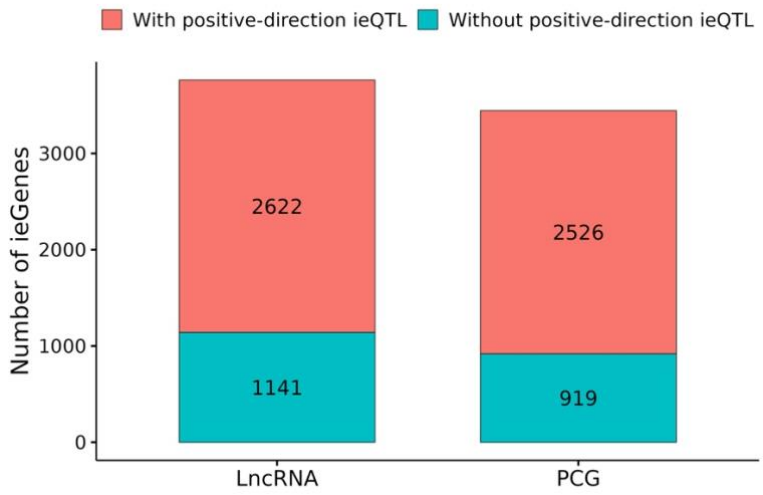

Figure R16. Number of ieGenes with or without positive-direction ieQTLs. Stacked bar plot showing the numbers of ieLncRNAs and iePCGs with or without at least one positive-direction ieQTL. Numbers within bars indicate ieGene counts.

Updated figure: Supplementary Figure 10

Updated table: Supplementary Table 4; Supplementary Table 5
Updated texts
(lines 180-181): "We retained 3,763 interaction-effect lncRNAs (ieLncRNAs) and 3,445 interaction-effect PCGs (iePCGs), each supported by at least one positive- or negative-direction ieQTL (**Supplementary Fig. 10**)."
(lines 304-306): "Of the lncRNA-ieQTL-supported colocalization events, 64.4% (76 out of 118) were supported by positive-direction ieQTLs (**Supplementary Table 4**)."
(lines 369-371): "By contrast, negative-direction ieQTLs showed weaker concordance with evidence supporting the inferred interacting cell type, indicating that their cellular attribution requires additional molecular evidence for interpretation."

13. The colocalization analyses use FDR < 0.25 for ieQTL selection, justified by reduced statistical power in interaction mapping. While this is precedented, it substantially increases the false positive rate among the input ieQTLs. The authors should provide a sensitivity analysis showing how colocalization results change under stricter thresholds (e.g., FDR < 0.05, FDR < 0.10), and explicitly discuss what proportion of the 142 colocalizing lncRNAs might be spurious.
Re: We agree with the reviewer that using a relatively relaxed threshold may increase the false-positive burden among the input signals for colocalization analysis. To address this concern, we now report the colocalization results using a more stringent ieQTL significance threshold of FDR < 0.05 in the revised manuscript (**lines 291-308**). Under this more stringent threshold, we identified 99 ieLncRNAs with colocalization evidence, of which 85 were absent from current GENCODE annotations. In the sensitivity analysis, we further compared colocalization results across multiple ieQTL thresholds, including FDR < 0.25, FDR < 0.10, FDR < 0.05, and FDR < 0.01. As expected, the absolute number of colocalized ieLncRNAs decreased as the ieQTL threshold became more stringent. However, the proportion of colocalized ieLncRNAs that were not annotated in GENCODE remained similar between the relaxed FDR < 0.25 threshold and the more stringent FDR < 0.01 threshold (**Figure R17 below**). This indicates that our conclusion that many colocalized ieLncRNAs are previously unannotated is not driven solely by the use of the relaxed FDR < 0.25 threshold.

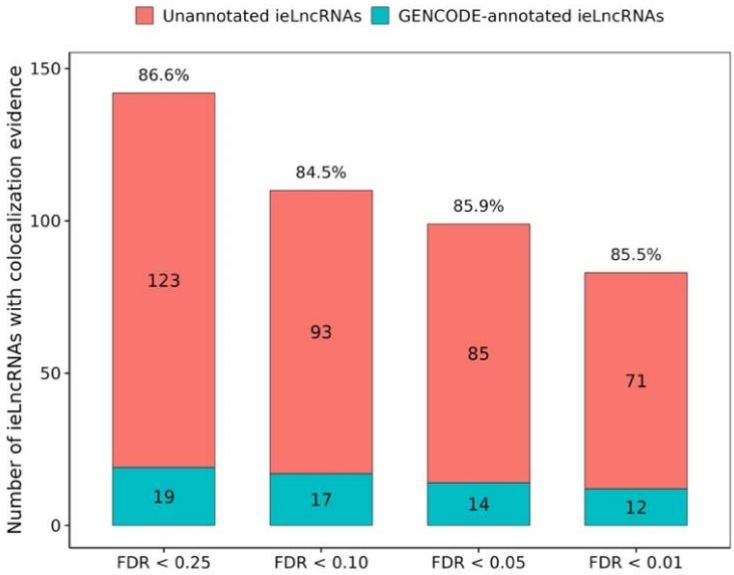

Figure R17. Colocalization results across different ieQTL significance thresholds. Stacked bar plot showing the numbers of unannotated and GENCODE-annotated ieLncRNAs with evidence of colocalization at four ieQTL significance thresholds (FDR < 0.25, 0.10, 0.05, and 0.01). Numbers within bars indicate the corresponding numbers of ieLncRNAs, and percentages above bars represent the proportion of unannotated ieLncRNAs among all colocalized ieLncRNAs at each threshold.

Updated table: Supplementary Table 4

Updated texts

(lines 291-308): "We used coloc[27] to investigate whether genetic associations for the brain-related traits described above may act through lncRNA-ieQTLs [5,6] (FDR < 0.05). For comparison, we also performed colocalization analyses using PCG-ieQTLs as well as matched lncRNA and PCG eQTLs. We identified 99 ieLncRNAs with evidence of colocalization (PP4 > 0.8), most frequently for traits such as schizophrenia, neuroticism, and educational attainment, whereas no colocalization signals were detected for multiple sclerosis or autism spectrum disorder (Supplementary Table 4). Notably, 85.8% (85 out of 99) of ieLncRNAs were absent from current GENCODE annotations, a pattern consistently observed under both more lenient and more stringent thresholds (**Supplementary Fig. 14**).

In total, 47.5% (77 out of 162) of lncRNA–trait colocalization events were uniquely identified through lncRNA-ieQTLs, whereas an additional 25.3% (41 out of 162) were supported by both lncRNA-ieQTLs and matched lncRNA eQTLs (Fig. 5a). Of the lncRNA-ieQTL-supported colocalization events, 64.4% (76 out of 118) were supported by positive-direction ieQTLs (**Supplementary Table 4**). These results underscore the added value of cellular-context-dependent regulatory mapping in detecting trait-relevant lncRNA loci and refining their likely cellular origins."

Minor comments

1. Line 85, full name of PCG is missing.

Re: Fixed (line 84).

2. Ref 37 and ref 38 are duplicated.
Re: We thank the reviewer for pointing this out. We have removed the duplicated reference and corrected the reference list accordingly.

3. Ref 64 is the Slingshot paper for trajectory inference. This appears to be a citation error for the qvalue package.
Re: We have corrected the reference information accordingly (Ref 68).

(Remarks on code availability)

Author well documented the code for ieQTL mapping with good annotations. But I did not see any code for lncRNA mapping.
Re: We have now added the code for lncRNA mapping to the repository, together with annotations to facilitate reuse.

References:

1. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat Genet* **51**, 187–195 (2019).
2. Sheng, X. *et al.* Mapping the genetic architecture of human traits to cell types in the kidney identifies mechanisms of disease and potential treatments. *Nat Genet* **53**, 1322–1333 (2021).
3. Tian, C. *et al.* Single-cell RNA sequencing of peripheral blood links cell-type-specific regulation of splicing to autoimmune and inflammatory diseases. *Nat Genet* **56**, 2739–2752 (2024).
4. Stephens, M. False discovery rates: a new deal. *Biostatistics* **18**, 275–294 (2017).
5. Kim-Hellmuth, S. *et al.* Cell type–specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020).
6. Kasela, S. *et al.* Interaction molecular QTL mapping discovers cellular and environmental modifiers of genetic regulatory effects. *The American Journal of Human Genetics* **111**, 133–149 (2024).
7. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**, e164 (2010).
8. Patrick, E. *et al.* Deconvolving the contributions of cell-type heterogeneity on cortical gene expression. *PLOS Computational Biology* **16**, e1008120 (2020).
9. von Bartheld, C. S., Bahney, J. & Herculano-Houzel, S. The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *Journal of Comparative Neurology* **524**, 3865–3895 (2016).
10. Herculano-Houzel, S. The human brain in numbers: a linearly scaled-up primate brain. *Front. Hum. Neurosci.* **3**, (2009).
11. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* **82**, 239–259 (1991).

12. de Klein, N. *et al.* Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. *Nat Genet* **55**, 377–388 (2023).

13. Dong, P. *et al.* Population-level variation in enhancer expression identifies disease mechanisms in the human brain. *Nat Genet* **54**, 1493–1503 (2022).

14. Kozlenkov, A. *et al.* A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. *Science Advances* **4**, eaau6190 (2018).

15. Mathys, H. *et al.* Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365-4385.e27 (2023).

16. Avila Cobos, F., Alquicira-Hernandez, J., Powell, J. E., Mestdagh, P. & De Preter, K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun* **11**, 5650 (2020).

17. Jew, B. *et al.* Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat Commun* **11**, 1971 (2020).

18. Sturm, G. *et al.* Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* **35**, i436–i445 (2019).

19. Chu, T., Wang, Z., Pe'er, D. & Danko, C. G. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat Cancer* **3**, 505–517 (2022).

20. Taylor-Weiner, A. *et al.* Scaling computational genomics to millions of individuals with GPUs. *Genome Biology* **20**, 228 (2019).

21. Chen, L. *et al.* Genetic control of non-coding RNAs in the human brain and their implications for complex traits. *The American Journal of Human Genetics* **0**, (2025).

22. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proceedings of the*

National Academy of Sciences **100**, 9440–9445 (2003).

23. Fujita, M. *et al.* Cell subtype-specific effects of genetic variation in the Alzheimer's disease brain. *Nat Genet* **56**, 605–614 (2024).

24. Fullard, J. F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome Res.* **28**, 1243–1252 (2018).

25. Bryois, J. *et al.* Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat Neurosci* **25**, 1104–1112 (2022).

26. Zeng, B. *et al.* Single-Nucleus Atlas of Cell-Type Specific Genetic Regulation in the Human Brain. 2024.11.02.24316590 Preprint at https://doi.org/10.1101/2024.11.02.24316590 (2024).

27. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genetics* **10**, e1004383 (2014).

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The authors have adequately addressed the concerns I raised in my report.
Re: We sincerely thank the reviewer for the positive feedback.

Reviewer #2 (Remarks to the Author):

The author has addressed my questions. Following minor issues need to be corrected.
Re: We sincerely thank the reviewer for the positive comments and for the constructive suggestions. We have carefully addressed all the remaining minor issues in the revised manuscript.

1. EPIGABA is listed as syn12034263 in Methods (line 449) but syn4588488 in Data Availability (line 630).
Re: Thank you for pointing this out. We have corrected it.

2. Reference 51 (Ghafouri-Fard et al.) is marked "RETRACTED" in the reference list but is still cited in the main text (line 280) to support the functional relevance of MIR124-3. This citation should be replaced with a non-retracted source.
Re: The reference has been corrected.

3. Methods (line 467–468) cites "Emani et al.45" for cell-type-specific chromatin accessibility peaks, but reference 45 is Nott et al.
Re: We have corrected the reference citation.

Reviewer #2 (Remarks on code availability):

I was able to install and run the code following the instructions provided in the README file.
Re: Thank you for the confirmation.